DOCUMENT RESUME

ED 369 821 TM 021 465

AUTHOR Harvey-Beavis, Adrian
TITLE Analyzing Teacher Judgments with the Rasch Model: Their Contribution to the Construction of Continua to Assess Literacy Developments in the Early Years of School.
PUB DATE Apr 94
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Behavior Patterns; Child Development; Educational Assessment; Elementary Education; Elementary School Students; *Elementary School Teachers; Evaluation Methods; Foreign Countries; Item Response Theory; *Literacy; Program Evaluation; Spelling; *Student Evaluation; Validity
IDENTIFIERS Australia; *Continuum Models; First Steps Program (Australia); Indicators; *Rasch Model; Validity Research

ABSTRACT

How teachers' judgments about student literacy behavior were analyzed under the Rasch model is described. The analysis was done to assist staff of the Western Australian Department of Education to revise aspects of a literacy program called "First Steps" for the early years of school. First Steps uses a developmental continuum of small phases of student development. To illustrate the analysis, the Spelling Development Continuum is used. Data for the validation of this continuum were provided by 39 teachers of years 1, 3, 5, and 7, with reevaluation data from an additional 88 teachers. Data consisted of judgments about the extent to which up to 10 students in each class exhibited evidence of behavior described by continuum indicators. Data were analyzed with the computer program Quest, which produces item- response-theory calibrations of indicators. The application of Rasch modeling techniques made it possible to extend the concern of the research beyond the order of the indicators to a more complex examination of the entire continuum, and this resulted in rewording of some indicators and more general investigations of the whole continuum. Nine exhibits demonstrate the analysis steps. (Contains 6 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
ADRIAN HARVEY-BEAVIS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED 369 821

Analysing Teacher Judgements with the Rasch Model:

Their contribution to the construction of continua to assess literacy development in the early years of school

Adrian Harvey-Beavis
Australian Council for Educational Research

Presented at the American Educational Research Association annual meeting
April 1994.

TM021465

I wish to thank the following people. Dr Ray Adams, Ms Jacqui Barley, Peter Congdon, Dr Geoff Masters and Ms Margaret Wu, all of ACER, for introducing me to Rasch modelling and guiding me through some its mysteries. I also wish to thank the staff at First Steps, and especially Ms Alison Dewsbury, for her assistance throughout the research project and for reviewing this paper. Any errors, in this paper, are mine.

The views expressed in this paper are not necessarily those of the Australian Council for Educational Research.

Introduction

The work reported here describes how teachers' judgements about student's literacy behaviour were analysed under the Rasch model. This analysis was done to assist staff of the Western Australian Department of Education revise aspects of a literacy program called 'First Steps'. First Steps is a project set up to assist teachers improve children's literacy in the early years of school.

The Australian Council for Educational Research (ACER) was commissioned by the Western Australian Department of Education to conduct the research reported here.

Some Background

First Steps

First Steps is a program instituted by the Western Australian Department of Education to improve the literacy of primary school students. (In Australia primary school children are generally aged between 5 and 12 years.)

First Steps assumes that literacy is acquired continuously in a developmental sequence. Literacy is seen, by First Steps, as involving four broad areas. For each of these areas the First Steps devised a continuum. The continua are; the 'Spelling Developmental Continuum', the 'Reading Developmental Continuum', the 'Writing Developmental Continuum', and the 'Oral Language Developmental Continuum'.

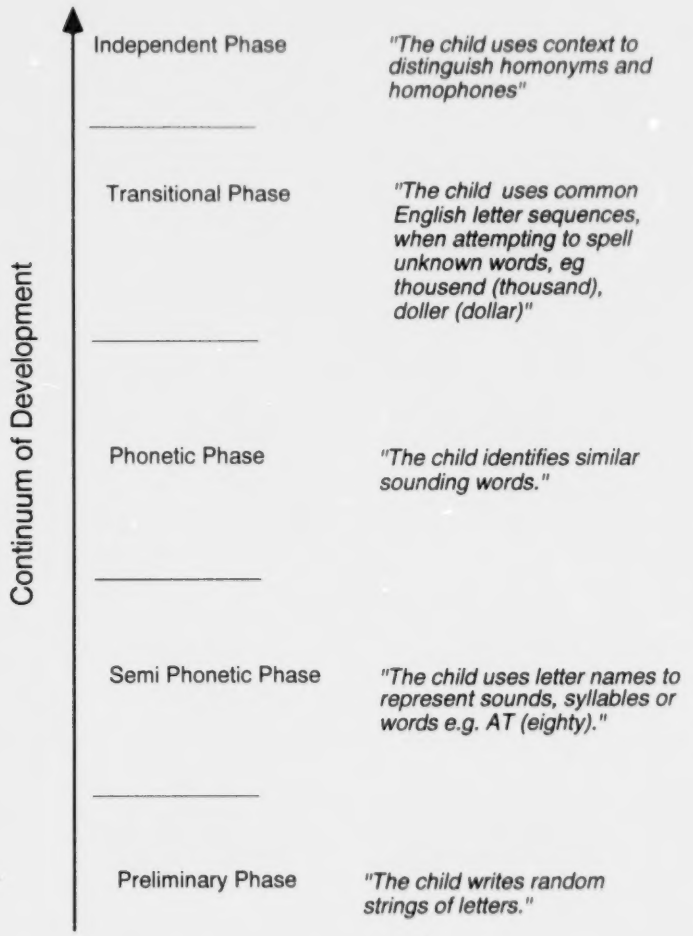
Each of these Continua consists of an ordered series of statements describing behaviours that children exhibit as they develop spelling, reading, writing and oral language skills. These behaviours are regarded by First Steps as like milestones marking out a child's development along the road to literacy.

During the construction and trialing of the Continua, it was observed that various indicators clustered together. These clusters of indicators were incorporated into the structure of the Continua and were named 'phases'. For example, the Spelling Developmental Continuum has five phases. These are named the 'Preliminary Phase', 'Semi Phonetic Phase', 'Phonetic Phase',

'Transitional Phase' and 'Independent Phase'. There are differing numbers of indicators in each phase. On average there are about 20 indicators per phase.

In summary, each First Steps Developmental Continuum consists of a small number of phases. Within each phase are statements (named 'indicators') that describe various behaviours associated with acquiring literacy. Exhibit 1 shows the phases, the order of their location along the continuum of development and an example indicator from each phase.

Exhibit 1: Overview of the First Steps Spelling Developmental Continuum


Continuum of Development
Independent Phase
"The child uses context to distinguish homonyms and homophones"
Transitional Phase
"The child uses common English letter sequences, when attempting to spell unknown words, eg thousend (thousand), doller (dollar)"
Phonetic Phase
"The child identifies similar sounding words."
Semi Phonetic Phase
"The child uses letter names to represent sounds, syllables or words e.g. AT (eighty)."
Preliminary Phase
"The child writes random strings of letters."

It is the phases that are used to describe the location of a child on a developmental continuum. After observing a child's behaviours, a teacher places a child into a phase. Once this is done, the most appropriate teaching

strategies can be used with the child. (First Steps provides many strategies, each linked to the various phases of development.)

From now on, only the Spelling Development Continuum is discussed. This was the one which was subjected to closest scrutiny during the ACER evaluation of First Steps.

Teachers judgements and the assessment of the First Steps Spelling Development Continuum

First Steps identified the behaviours incorporated into the Spelling Developmental Continuum by only using data from published research. The description of these behaviours was also taken from the research literature although these descriptions were modified during the development of First Steps. Examples of spelling behaviours, sometimes used to illustrate indicators, were only drawn from children's work samples. Generally, however, neither the indicators nor their descriptions were obtained from the classroom teachers of Western Australia. In this sense, the Spelling Continuum was devised from the 'top down' - from the published judgements of experts.

The next stage in the evolution of the Spelling Continuum was to establish the extent to which it correctly describes the behaviours exhibited by children as they learn to spell. The ACER was commissioned to undertake this stage. To do this, classroom teachers were asked to make judgements about whether students they were teaching did or did not exhibit the spelling behaviours described by the First Steps Spelling Developmental Continuum. These judgements formed the data for the empirical validation of the Spelling Continuum. The data were then analysed under the Rasch model. The results of this analysis were used to identify 'problem indicators'. These indicators were then modified and a new version of the Spelling Developmental Continuum created. This new version was then, also validated using the same methods.

In summary, the empirical validation of the First Steps Spelling Developmental Continuum used judgements made by teachers about the spelling behaviours of their students to assess the extent to which the Continuum accurately depicted the development of spelling in children. These teachers' judgements were,

therefore, the criterion by which the accuracy of the Spelling Continuum was assessed.

The Sample

The data for the validation of the Continuum were collected in November 1992. The data for the revalidation of the revised Continuum were collected in November 1993.

The data for the validation of the Spelling Continuum were provided by teachers of Years 1, 3, 5 and 7 students from a sample of Western Australian government schools. Altogether 50 teachers were asked to participate. Of these 50 teachers, 39 returned data that could be used. For the revalidation, 99 teachers were asked to participate and 88 returned useable data. No teachers who participated in the first validation were asked to participate in the revalidation. Exhibit 2 shows the numbers of teachers who responded for each Year level. It will be noted that in the revalidation Kindergarten teachers were included for the first time. This was done so that some indicators from the Preliminary Phase, not calibrated in the 1992 validation could be calibrated in 1993. These indicators were not calibrated because all students exhibited these behaviours. It was thought that not all kindergarten children would exhibit these behaviours.

Exhibit 2: Number of teachers providing data for the 1992 validation and the 1993 revalidation of the Spelling Developmental Continuum for each Year level.

	Kinder	Yr 1	Yr 3	Yr 5	Yr 7	Total
1992 Validation	0	10	8	12	9	39
1993 Revalidation	14	20	20	20	14	88

The Data

The data consisted of judgements made by teachers about the extent to which each of up to ten students in their class exhibited evidence of having demonstrated the behaviour described by each of a number of First Steps indicators. (The teachers made these judgements about each child in turn and not about the group of children.) These judgements were recorded on a computer used by the teachers. They tapped one of a set of appropriate keys to

register their response. There were two sets of responses available to a teacher depending upon which indicator was displayed to them. The response set to be used was pre-determined. Teachers could not choose which response set to use.

The first response set was:

Y - (Yes/Most of the time)
This key was to be pressed if the named student usually demonstrated the behaviour described by the indicator.
N -(No/Hardly Ever)
This key was to be pressed if the named student did not or hardly ever demonstrated the behaviour described by the indicator.
U - This key was to be pressed if the teacher was unable to make a judgement. If the teacher responded with 'U', the program asked the teacher to select one of the following:

1 - "I have not yet had an opportunity to gather information relating to this indicator."
2 - "I don't understand the wording of this indicator."

If the teacher had understood the indicator, then after all students had been evaluated with respect to that indicator, the results of the teacher's judgements were displayed on the screen and an opportunity was given to alter the data.

The use of the first response set was not appropriate for all indicators. If only this response set was used, then it would have led to ambiguous responses for some indicators. A response 'Hardly ever' could mean that the child has yet to learn a particular behaviour described by an indicator. It could also mean that the child has advanced beyond this behaviour and so abandoned it (An example of one such indicator is: 'The child writes random strings of letters'. A child may 'hardly ever 'exhibit this behaviour because it is yet to learn how to write words. A child may also 'hardly ever' exhibit this behaviour because it no longer writes letters randomly.) To overcome this problem, a second response set was developed. This second response set had the additional category of 'Beyond'.

The second response set was:

B - (Beyond this level.)
This key was to be pressed if (1) the indicator was phrased negatively and the student demonstrated evidence of being able to perform the converse

of the indicator, (2) the indicator began with the phrase "Beginning to ..." and the student had fully acquired the skill referred to in to the indicator or (3) a student had developmentally 'left behind' the indicator and so no longer demonstrated evidence of it (as opposed to having yet to develop this skill).

Y - (Yes/Most of the time)
This key was to be pressed if the student usually demonstrated the behaviour described by the indicator.

N -(No/Hardly Ever)
This key was to be pressed if the student did not or hardly ever demonstrated the behaviour described by the indicator.

U - This key was to be pressed if the teacher was unable to make a judgement. (The same categories - 1 or 2 - as for the first response set were then displayed.)

In the data analysis 'Yes' and 'Beyond' were treated as having identical meaning and so coded to the same value. (Responses with the value 'U1' and 'U2' were excluded from the analysis.

Most teachers were asked to provide data for the first five female students on a class list and for the first five male students on a class list. If a class had fewer than five female students or fewer than five male students, then teachers were asked to 'top up' with other students from the class. In some small schools and in some composite classes there were less than ten students at a given Year level. In these cases teachers were asked not to top up with students from other Year levels. Exhibit 3 shows the total number of students involved in the study for each Year level.

Exhibit 3: Number of students involved in the study for the 1992 validation and 1993 revalidation of the Spelling Developmental Continuum for each Year level.

	Kinder	Yr 1	Yr 3	Yr 5	Yr 7	Total
Validation	0	90	74	106	83	353
Revalidation	136	197	191	174	145	843

It was decided to design the data collection in such a way as to avoid asking teachers about indicators which would be unlikely to be observed in their students. For example, Kindergarten and Year 1 teachers were not asked to provide data about the indicators in the Transitional or Independent phases because it was felt to be most unlikely that any Kindergarten or Year 1 students would have developed such high levels of spelling skill. To ensure that the data could still be used to depict a sequence of development across all phases of the Continuum, each Year level had at least one phase in common with the Year level below it or with the Year level above it. Exhibit 4 shows which phases of the Spelling Developmental Continuum were used by Year level, how these phases overlapped and how many indicators were in each phase. The ticks in Exhibits 4 show which phases were used with each Year level. For example, in Exhibit 4, the column under the title 'Yr 1' indicates that Year 1 teachers provided data on indicators drawn from the 'Preliminary', the 'Semi Phonetic' and the 'Phonetic phases'. Exhibit 5 shows the design for the 1993 revalidation study.

Exhibit 4: Number of indicators per phase of the First Steps Spelling Developmental Continuum for each Year level and distribution of phases across Year levels for the 1992 validation.

Spelling D. C. Phases	N. of Indicators	Yr 1	Yr 3	Yr 5	Yr 7
Independent	20				√
Transitional	20		√	√	√
Phonetic	23	√	√	√	√
Semi Phonetic	21	√			
Preliminary	24	√			
Total N of Indicators	108	68	43	43	63

In this way, teachers were allocated phases according to the Year level that they taught. The phases were matched to Year level using advice from the First Steps project personnel. The matching was designed to ensure that the chosen phases were appropriate to the level of development of the students. Inappropriate phases were, either, ones that contained indicators which all students of a given

Year level would exhibit, or which contained indicators which no students, of a given Year level, would exhibit. A mix of student competencies was needed.

Exhibit 5: Number of indicators per phase of the First Steps Spelling Developmental Continuum for each Year level and distribution of phases across Year levels for the 1993 revalidation.

Spelling Continuum Phases	N. of Indicators	Kinder	Yr 1	Yr 3	Yr 5	Yr 7
Independent	20					√
Transitional	19			√	√	√
Phonetic	23	√	√	√	√	√
Semi Phonetic	21	√	√			
Preliminary	23	√	√			
Total N of Indicators	106	67	67	42	42	62

Method of Data Collection

All teachers were sent a computer disk. On this disk was a computer program written by staff at ACER. When the program was run, it prompted teachers for responses to questions.

Teachers were first asked the following questions about each child:

- What is the child's sex? (M/F)
- Is English the first language of the child? (Y/N)
- Is the child an Aboriginal or Torres Straits islander? (Y/N)
- Is the child receiving English as a Second Language assistance? (Y/N)
- Does the child have a disability that could significantly affect achievement in English? (Y/N)

Once this was done, the teacher was presented the text of an indicator with the instruction to assess each student with respect to the indicator on display. The indicators were presented to the teachers in a random order. The teachers knew only that the indicators came from the Spelling Developmental Continuum and that the indicators came from one of two or three phases. They did not know, unless they recalled it from their own use of the Continuum, from which phase a displayed indicator came.

These responses were stored on the disk and when the teacher had entered all the data, the disks were returned to ACER.

It was estimated that teachers would take approximately 2 hours to make all their judgements and enter the data. The computer program was designed so that teachers could quit before completing all the data entry and resume later.

How the data were analysed

The data for the 1992 and 1993 empirical validations of the First Steps Spelling Developmental Continuum were analysed using the computer program *Quest* (Adams and Khoo, 1992), which produces Item Response theory calibrations of indicators.

The data analysis

The main aim of the data analysis was to 'calibrate' the indicators for the Spelling Developmental Continuum. The calibration process estimates a 'difficulty' level for each indicator. In general, the greater the number of students displaying the behaviour described by an indicator, the 'easier' that indicator is estimated to be. In this way, the calibration process parallels the intention of First Steps by seeking to locate indicators at positions along a Continuum.

For the purposes of the validation, it was assumed that all children were likely, at some stage of their development, to exhibit the behaviour described by an indicator. If an indicator, no matter how intrinsically easy it is, was never exhibited by a large proportion of children it would be estimated to be difficult under the Rasch model. The assumption that most children would at some time exhibit these behaviours was judged to be reasonable. First Steps personnel claimed that most children would exhibit all the behaviours described by the indicators on the Spelling Developmental Continuum.

There is another important assumption underlying the analysis. It was assumed that there is only one underlying continuum that goes to make up the process of learning to spell. It could be argued that learning to spell is a multi dimensional

process which cannot, therefore, be captured using a single continuum. The judgement made, however, was that spelling development can be captured using a single continuum. This judgement was based upon the view that it is reasonable to claim that some children spell better than others. In such a claim there is an implicit assumption that there is only one organising dimension being tapped so the use of one continuum is appropriate.

Results

Exhibit 6 shows the estimated difficulty of each indicator grouped by phases. To complement this exhibit, another display, Exhibit 7, was constructed to reduce the effect of outlying estimates of difficulty.

Exhibit 7 uses the median to represent the measure of central tendency and the interquartile range to represent the spread. These measures are more resistant to the effect of outliers than the mean and the standard deviation. The data in Exhibit 7 are displayed using a type of box plot. The lower boundary of the box identifies the value above which 75% of the estimates fall and the upper boundary of the box identifies the boundary below which 75% of the estimates fall. Thus, 50% of the estimates are located between the top and the bottom of the box. The horizontal line inside the box marks the location of the median. In some boxes the median is not located centrally. In these cases the data are concentrated on the side of the box that is closer to the median. (See, for example, the Phonetic Phase in Exhibit 7. Here the data are concentrated towards the top end of the box.) The horizontal width of a box is a function of the number of indicators in the phase. The more indicators there are in a phase, the wider is the box.

Exhibit 6: Difficulty estimates of indicators within phases of the First Steps Spelling Developmental Continuum.

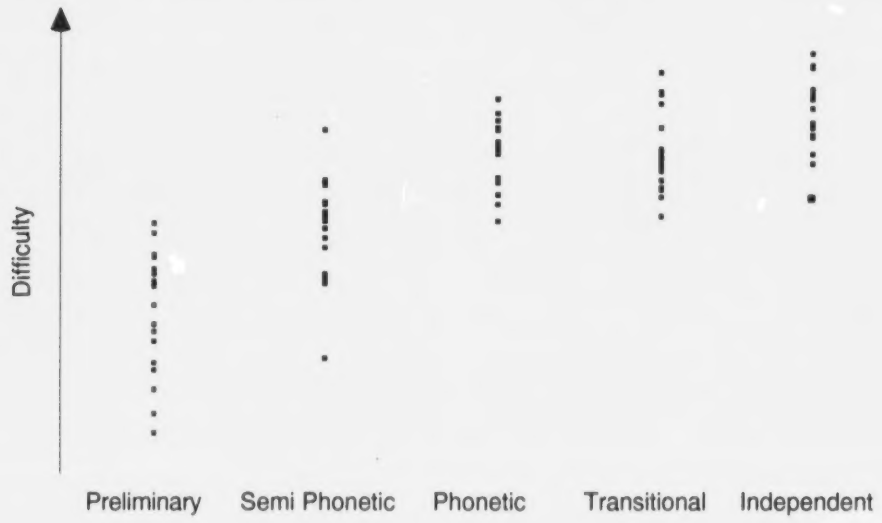
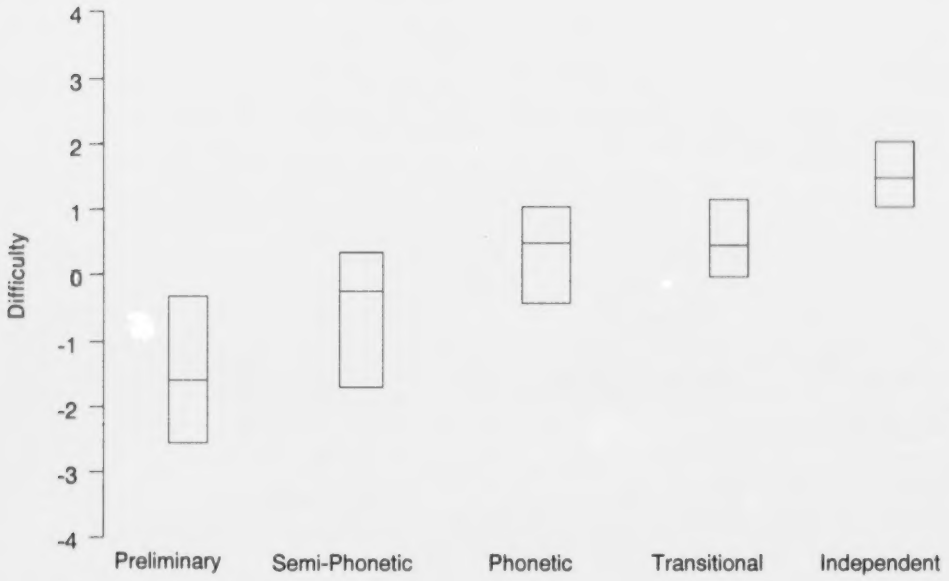
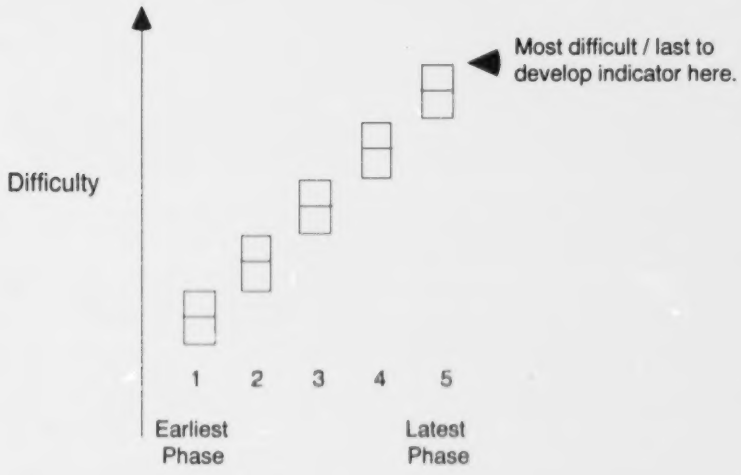


Exhibit 7: Box plots of difficulty estimates of indicators within phases of the 1992 version of the First Steps Spelling Developmental Continuum.



Once the data had been displayed, the question at the very heart of the validation arose - to what extent has First Steps got the order of the indicators correct? To help answer this question, an ideal model was proposed. This ideal model is displayed in Exhibit 8.

Exhibit 8: An ideal model showing the distribution of indicators of a developmental continuum when ordered by phases.



The height of the boxes in Exhibit 8 represents the spread of difficulty of the indicators within each phase. The main features of this model are:

(1) the phase depicting the earliest stages of development has all of its indicators situated towards the bottom or less difficult end of the scale,
(2) the phase depicting the most advanced stages of development has all its indicators situated towards the top or most difficult end of the scale,
(3) the other phases fall between these extremes in the appropriate order,
(4) the spread for each phase along the scale of difficulty is the same
(5) there is no overlap between the phases,
(6) there is no gap between the phases,

Early phases should contain indicators low on the difficulty scale, intermediate phases should contain indicators of intermediate difficulty and later phases should contain indicators of highest difficulty. If the phases are not organised in this way, then they are not depicting the sequence of development that children

go through. The reason for having the phases is to depict this sequence of development.

There should be a spread of difficulty estimates within a phase because development is seen as occurring along a Continuum and not in stages. If development occurred in stages it would be expected that each indicator within a phase would have the same estimated difficulty. However, the spread of indicators within a phase should not be too wide because this will lead to poor discrimination when plotting the development of children. There is little point in allocating a child to a stage if that stage, for example, covers a significant span of their school years. Reasonably fine levels of discrimination are required if development is to be charted and the indicators within a phase are to operate as something more than a checklist of skills. This can be achieved by having the phases ordered along the difficulty scale such that each occupies the same proportion of the total spread of the scale. It should be noted, however, that while 'equal spread' might be thought of as an ideal, it is not necessary to the successful construction and use of a Continuum.

The spread of the estimates of difficulty within one phase should not overlap with the spread in any other phase because this can lead to difficulty in establishing the level of development of the child. For example, take the extreme case where the spread of estimated difficulty of indicators within two phases entirely overlaps the other. When this occurs, allocating a child to one of those phases does not assist in locating that child along the developmental Continuum. This suggests that the more overlap there is between phases, the more ambiguity there will be about the location of the child on the Continuum.

Ideally, gaps between the spread of estimates of difficulty within one phase and an adjoining phase should not occur. A gap means that if the child is at a location in their development along the Continuum where this gap occurs, then their level of development may be under-estimated by a teacher using such a Continuum. However, as a gap does not lead to a confusion about the sequence of development of the child along a Continuum, it is less of a problem than having large overlaps between the phases.

The ideal type represents the continuum as it should exist in its ideal form. If the experts' opinions, expressed by their published research, are correct about which behaviours are pertinent and if they are ordered appropriately by First Steps, then

this ideal type should be reproduced by using the calibrated judgements of teachers.

When the calibrations for the 1992 validation were examined, it was judged that the spread of difficulty of the indicators in the Preliminary Phase was rather long and the Independent Phase somewhat short when compared with the ideal model. Overall, however, the spread of indicators within phases was acceptably. It was also judged that there was excessive overlap between most of the phases. This was judged to be a concern. There were no gaps between the phases.

There was, clearly, room for improvement in the broad patterns seen in Exhibit 7 and in specific indicators. The examination of the individual indicators was especially important for the revalidation.

First Steps and ACER staff worked together to examine individual indicators. This was done with to a view to either changing the wording or deleting 'problem' indicators from the Spelling Continuum. This was done so that the Continuum would more accurately and precisely describe the order of behaviours exhibited by children as they learn to spell.

Three approaches were used to evaluate individual indicators. First, the proportion of teachers who did not understand each indicator was examined. Secondly, the estimated difficulty of each indicator was examined. The wording of these indicators with estimates of difficulty above the 90th percentile and those below the 10th percentile within the Semi Phonetic, Phonetic and Transitional Phases was examined. Indicators above the 90 percentile of the Preliminary phase were also examined, as were indicators below the 10th percentile of the Independent phase. (It was assumed that no one indicator within the earliest phase of development could be too easy and that no one indicator in the Independent phase could be too difficult.) Thirdly, the wording of indicators with a poor Infit Mean Square was examined.

Nearly all teachers understood all the Spelling Continuum indicators. Only one change was made to an indicator because it was not understood. (Note that all teachers had received considerable training in the use of First Steps. This, therefore, was not a surprising finding.)

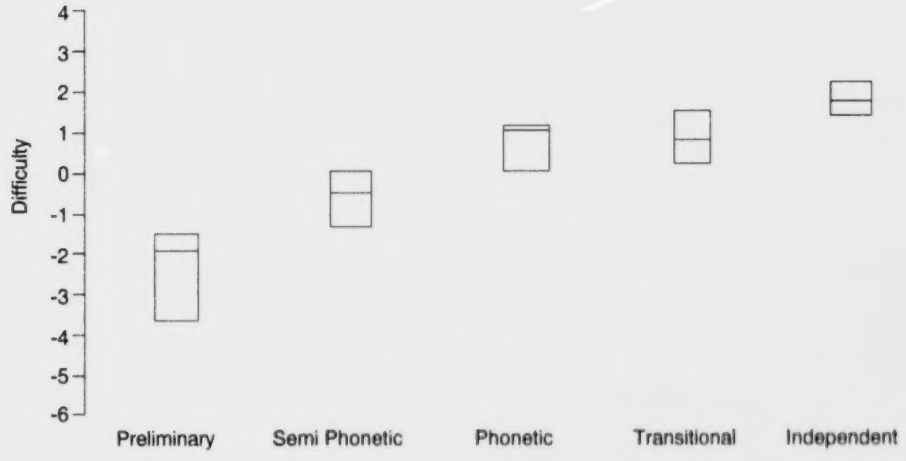
In total, 18 indicators were identified as having an extreme estimate of difficulty when compared with the other indicators in their phase. Of these 18, 2 were omitted from the continuum, 7 had their wording changed and 9 were left

unchanged. The unchanged indicators mostly came from the Preliminary Phase. It was felt that any changes made to these indicators should await the inclusion of Kindergarten teachers in the revalidation.

There were 15 indicators which were identified with a poor Infit Mean Square. (A poor Infit Mean Square - less than about 1.25 - was interpreted as meaning that teachers were not consistently interpreting these indicators or, alternatively, that the indicators were not tapping spelling ability.) Of these 15 indicators, 2 were omitted from the Continuum, 10 were changed and 3 were left unchanged.

Once these changes were made, the new version of the Spelling Developmental Continuum was revalidated. The results of the revalidation are shown in Exhibit 9.

Exhibit 9: Box plots of difficulty estimates of indicators within phases of the 1993 version of the First Steps Spelling Developmental Continuum



It shows that there is a clear improvement in the Spelling Developmental Continuum. Generally, each successive developmental phase groups indicators into bands which reflect increasing difficulty. There remains, however, some levelling out of the slope between the Phonetic and Transitional Phases. The boxes in Exhibit 9 are approximately the same length, although the Preliminary Phase is still rather longer than the others and the Independent Phase is still somewhat shorter. Most phases thus have an appropriate width to their spread of estimated difficulty. The observed data, in this regard, come close to matching the ideal model. More importantly, there is now very little overlap between any of

the phases. The exception is the Transitional Phase which still considerably overlaps with the Phonetic Phase. The earlier version of the Spelling Continuum had a large overlap between the Preliminary and Semi Phonetic Phases, a noticeable overlap between the Semi Phonetic and Phonetic Phases as well as a large overlap between the Phonetic and Transitional Phases. In the revised version the overlap between the Phonetic and Transitional Phases is all that remains. In this respect, then, the modifications made to the Spelling Continuum have led to a clearer and more precise depiction of the order of the behaviours children exhibit as they develop their spelling skills.

The revised version also has greater coherence in that the number of indicators with a poor Infit Mean Square was reduced from 15 to 7. As well, all indicators were reported by teachers to be well understood.

Reflection upon the use of IRT and the validation of the First Steps Spelling Developmental Continuum

The processes of revising the Spelling Continuum using the results of the empirical validation were interesting for the way in which they consistently highlighted weaknesses in various aspects of the original version of the Continuum.

Some of the indicators which were highlighted as a problem were worded negatively. An indicator is a behaviour. The notion of a negative behaviour seems to be self evidently confused and therefore to be avoided. The validation highlighted this problem.

Another problem which the validation highlighted was the importance of the exemplars that were used to illustrate an indicator. For example, one indicator from the Semi Phonetic phase was originally worded "The child writes one or two letters for sounds and then adds random letters to complete the world *eg crecuea (creature).*" This indicator was estimated under the Rasch model to be too difficult for the phase in which it was located. An examination of the exemplar showed that this was not an example from the semi phonetic phase but from the developmentally later Phonetic phase. The indicator was changed to read "The child writes one or two letters for sounds and then adds random letters to complete the world *eg greim (grass), rdms (radio).*" This reworded indicator

does not have an extreme estimate of difficulty in the new version of the Continuum.

Other clarifications also occurred. For example, one indicator from the Semi Phonetic phase was 'The child relies heavily on the most obvious sounds in a word ..' This was estimated under the Rasch model to be too difficult. It was reworded to 'The child relies on the sounds which are most obvious to him or her. This may be the initial sound, initial and final sounds or initial, medial and final sounds ...' Not only is there now more precision in this indicator, but, importantly, the teacher is now being cued not to presume that the sound which they (the teacher) hear as the most important will be perceived by the child as the most important.

Perhaps the most powerful outcome of the validation was in terms of the overall coherence of the Spelling Developmental Continuum. Not only was it shown to be reasonably cohesive, and with the changes made after 1992, more cohesive, but it raised questions about the structure of the Continuum. For example, there was, and remains, the problem with the Transitional phase. No satisfactory explanation has yet been found for why this phase seems, on average, to be no more difficult than the earlier Phonetic phase. It has become clear, as this problem has been thought about by First Steps personnel, that the idea of a 'Transitional phase' is itself problematic. The other phases are named to indicate a general level of spelling achievement. To be 'in transition' tells nothing of the contents of that general level of achievement described by this phase. This has set in train thinking about what it is that characterises this phase of development and so, at the most general level, how the learning to spell is to be properly understood.

Conclusion

The empirical validation of the First Steps Spelling Developmental Continuum began with an apparently simple concern with the order of the indicators which made up this Continuum. Item response theory - in particular the application of Rasch modelling techniques - enabled this initial concern to lead to a more complex interrogation of the Continuum. In particular it led not only to the rewording of individual indicators increasing their clarity, but also to more general investigations into the structure of the Developmental Continuum as a whole.

Bibliography

Adams, R. and Khoo, S., *Quest*, Hawthorn, ACER, 1992

Western Australian Ministry of Education, *Spelling Developmental Continuum*, Perth, 1992

Western Australian Ministry of Education, *Reading Developmental Continuum*, Perth, 1992

Western Australian Ministry of Education, *Writing Developmental Continuum*, Perth, 1992

Western Australian Ministry of Education, *Oral Language Developmental Continuum*, Perth, 1992

Wright, B. and Stone, M. *Best Test Design*, Chicago, Mesa Press, 1979.